

# **Developing a Performance Evaluation System for Charter Schools: The Balanced Scorecard Approach**

Richard S. Brown, Ph. D.  
Priscilla Wohlstetter, Ph.D.

Center on Educational Governance  
Rossier School of Education  
University of Southern California

## **Introduction and Purpose**

Federal pressure on states and districts to hold public schools accountable for student performance is not new. A system of standards, assessments, and accountability were central to the Elementary and Secondary Education Act under the Clinton Administration (1992-2000). However, in 2001, Congress added some teeth to the system when it enacted the No Child Left Behind Act (NCLB). NCLB forces states to refine their accountability systems to include tracking student performance by certain grades, in certain disciplines, and by student subgroups. The federal law also requires state accountability systems to include targets for progress to ensure that all schools and districts make Adequate Yearly Progress (AYP), based on assessments included in the statewide accountability system. NCLB also added consequences for poor performance, if a school fails to achieve AYP in two consecutive years, it is placed in “improvement status” (Debray, McDermott, & Wohlstetter, 2005) and faces potential restructuring, closure and reconstitution.

This article describes a process of developing a set of performance indices for charter schools using data that public schools routinely report to state educational agencies for compliance purposes. Charter schools are public schools funded by the state but granted greater flexibility than other public schools in return for being held accountable for their students’ educational progress. At the end of the charter contract period, usually 3 or 5 years, the charter may be revoked if the school has not successfully accomplished its performance outcomes. While the need for performance information is especially important to the survival of charter schools, the resources available for data collection and analysis are weak. Many charters are small, have few administrators, and have fewer dollars per student than district-run public schools. Consequently, charter schools generally do not have the resources to develop on their own multiple performance measures and to identify promising practices that would improve student achievement.

Every year data elements submitted by public schools to the state in response to accountability requirements, specifically student achievement data, are released by the state, districts and schools to much public attention. Local and statewide newspapers publish complete lists of these scores, press conferences are held by the state and district superintendents, and school principals

and teachers hold community and parent meetings to provide their own evaluations of student and school performance. So what do these data tell us? The data tell the public how well schools are meeting established educational goals. However, the data do not provide detail or analysis to determine what areas may be limiting educational performance.

In California, the Public School Accountability Act (PSAA) calls for an Academic Performance Index (API) that includes multiple indices beyond test scores, such as school staff attendance and student graduation rates. In fact, California requires all public schools to submit a School Accountability Report Card (SARC) that includes data on the demographics of students, teachers, and staff; school safety; learning climate; academic data; school completion; class size; curriculum and instruction; postsecondary preparation; and fiscal and expenditure data. However, California has only integrated test score results into the API to date. The purpose of this article is to describe the principles, process, and lessons learned of creating a performance evaluation system for charter schools that focuses on the development of multiple measures that reflect the various dimensions of student learning, program effectiveness, and school operations within charter schools.

## **Framework**

The framework for the performance evaluation system was based on the idea of “the balanced scorecard” (Kaplan & Norton, 1992, 1993; Meyer, 1994). The balanced scorecard is a way of looking at multiple measures of a system’s performance in a balanced way, rather than focusing exclusively on a single indicator, which is specifically relevant in educational systems where the evaluative focus is primarily, if not exclusively, on student test scores. The balanced scorecard approach assists organizations in developing and focusing on measurable goals. Kaplan and Norton identify four key perspectives that a balanced scorecard approach to measuring a system’s performance should include. These are: a customer’s perspective, an internal perspective, a financial perspective, and a learning or innovation perspective.

Although originally developed for application in the commercial environment as a means of measuring performance in ways that extended beyond merely financial indices, it is nonetheless applicable and has been deployed in other settings. The balanced scorecard approach has been used to evaluate system performance in such diverse settings such as hospitals (Pink, et al, 2001), health care management organizations (Urrutia & Eriksen, 2005), and local governments (Quinlivan, (2002). In addition it has recently been applied in higher education settings (Storey, 2002), specifically in measuring the performance of these systems with respect to equity and diversity issues (Bensimon, 2004, 2005).

Consistent with perspective proposed here, earlier research on public school accountability reporting suggests that when developing school accountability reports, it is important to present “the multivariate nature of the school” rather than focus solely on a few indices like test scores (Brown, 1999). Thus, we created a system of indices for charter schools that provides a more comprehensive and balanced approach to performance evaluation. In doing so, we delineated four principles for designing a performance evaluation system for school accountability using the balanced scorecard approach. The development process must utilize multiple measures, use

publicly available data, involve potential users in the development, and, inform diverse audiences.

## **Methodology and Data Sources for Developing the Indices**

To this end, a series of such measures was created. Charter School Indices -USC (CSI-USC) is a comprehensive, quantitative database that allows comparisons of California charter schools with other public schools on multiple measures of school, staff, and student performance. CSI-USC uses academic and financial data generated by the existing school accountability system for the state and refines it into a workable set of indices to measure various dimensions of school performance. All the data for the CSI-USC indices are drawn from required reports filed annually by California charter and other public schools. CSI-USC gathers and calculates these indices for *all* California charter schools each year and reports annually on how charter schools compare to other public schools. In a few cases where financial data are not available for all public schools, CSI-USC reports year-over-year trends for California charter schools. So, for a given year, we are able to examine how *any one* California charter school performed on a given measure and, at the same time, how that school compared to *all* other California charter schools on that measure and to non-charter public schools as a group.

Currently CSI-USC has 11 indices grouped into four performance categories: 1) financial resources and investment; 2) school quality; 3) student performance; and 4) academic productivity. The indices use a standardized rating system (1-10) to provide a comprehensive picture of charter school performance. Further, to include the other perspectives as identified by Kaplan and Norton (1993) CSI-USC supplements the indices with stakeholder satisfaction surveys of parents, students, and teachers.

A critical component to determining the scientific soundness of each performance index is the process utilized in scale development. For most of the indices, it was necessary to combine data elements for each of the indices representing that index to form a valid and reliable scale. In most cases, these index scales were developed using standard principal components factor analytic techniques with scale scores calculated using a regression method. That is, where applicable all of the indicator variables representing a given index were factor analyzed and factor scores were created using accepted and appropriate techniques. In all cases, the factor analyses resulted in a single factor explaining the vast majority of the variance among the indices, providing validity and reliability evidence of the constructs of interest.

Once factor scores were created, they were rank ordered into deciles to provide easy to understand index values ranging from 1 to 10 on each index. In most cases, these decile rankings were established using full data from the public school data files. In other cases, such as with the Financial Health Index, the decile rankings were predicated on the available data which represented charter schools only. Below is a description of the 11 indices organized by the four performance categories.

## Financial Resources and Investment

- 1. Financial Health Index:** A charter school with a high level of financial health has a stable level of cash on hand, as measured by its liquidity and reserves ratio. Liquidity is the ratio of assets to liabilities for a given school's financial statement and reserve ratio is the ratio of reserve fund balances to revenues. Both measures are drawn from the financial accountability form submitted annually by each charter schools to the state department of education. These variables are combined using a linear combination in a way that maximizes the variance accounted for between schools.
- 2. Direct Classroom Investment Index:** A charter school with a high direct classroom investment rating invests a significant portion of its financial resources in classrooms as distinct from applications outside of classrooms. This index is derived as the ratio of classroom investment relative to total revenues. The total revenues number is drawn from the financial accountability form submitted annually by each charter school to the state department of education. The classroom investment number represents expenditure categories such as teachers' salaries.

## School Quality

- 3. Learning Environment Index:** A charter school with a high learning environment rating has low pupil-staff and pupil-teacher ratios (relatively high proportions of adults working with students). These two ratios are calculated by taking the number of staff and the number of teachers from the School Accountability Report Card (SARC) data files, respectively, and dividing by the number of students enrolled in the school (also obtained from the SARC data files). These variables are combined using a linear combination in a way that maximizes the variance accounted for between schools.
- 4. Teacher Qualification Index:** A charter school with a high teacher qualification rating has a team of teachers with relatively more credentials and experience. This index was computed in accordance with Ken Futernick's original formulation [see [www.edfordemocracy.org](http://www.edfordemocracy.org) for derivation of measure) using the percentage of teachers on emergency, intern, or waiver credential as well as the percentage of teachers who are in their first or second year of teaching. The data are drawn from the SARC data files.
- 5. Reclassification Index:** A charter school with a high reclassification rating is integrating its English learners into the general education system at a higher rate than is a charter school with a low reclassification rating. This index is computed as the ratio of two measures, the number of students reclassified as Fully English Proficient and the number of students in the prior year who were English-Language Learners. These data are drawn from the SARC data files.

## Student Performance

- 6. Academic Performance Index (API):** The API measures the academic performance and growth of schools. A school's score on the API is an indicator of a school's performance

level. A school's growth is measured by how well it is moving toward or past that goal. This index is constructed using the API school rank, API similar schools rank, and API base score. These variables are combined using a linear combination in a way that maximizes the variance accounted for between schools.

- 7. Adequate Yearly Progress (AYP):** Under AYP criteria, California schools and numerically significant student subgroups must: 1) meet Annual Measurable Objectives (AMOs) in English/language arts (ELA) and mathematics; 2) demonstrate a 95-percent participation rate on assessments in ELA and mathematics; 3) demonstrate progress on the Academic Performance Index (API); and, 4) demonstrate progress on the graduation rate of its students (high school only). This index is constructed using four measures (met AYP in math, met AYP in ELA, percent proficient or above in math, percent proficient or above in ELA). These variables are combined using a linear combination in a way that maximizes the variance accounted for between schools.
- 8. Academic Momentum Index (AMI):** A charter school with a high academic momentum rating is improving student achievement over time. This index is constructed using three measures of academic progress (annual change in the proficient or above in math, annual change in proficient or above in ELA, and annual API growth). These variables are combined using a linear combination in a way that maximizes the variance accounted for between schools.

## **Academic Productivity**

- 9. English/Language Arts (ELA) Productivity Index:** A charter school with a high ELA productivity rating has a higher percentage of students proficient in ELA and lower expenses than similar schools. This index is constructed by comparing schools with similar funding levels on the percent proficient or above in ELA on the California Standards Tests. Funding levels for most schools are defined as the district average ADA, but for charter schools the level is calculated based on revenues and enrollments.
- 10. Math Productivity Index:** A charter school with a high math productivity rating has a higher percentage of students proficient in math and lower expenses than similar schools. This index is constructed by comparing schools with similar funding levels on the percent proficient or above in math on the California Standards Tests. Funding levels for most schools are defined as the district average ADA, but for charter schools the level is calculated based on revenues and enrollments.
- 11. Overall School Productivity Index:** A charter school with a high overall school productivity rating has a higher API score and lower expenses than similar schools. This index is constructed by comparing schools with similar funding levels on base API scores. Funding levels for most schools are defined as the district average ADA, but for charter schools the level is calculated based on revenues and enrollments.

## **Results and Lessons Learned**

The process of selecting, calculating, and validating meaningful indices of performance yielded numerous lessons. These include:

1. Be flexible in developing measures; compromises are inevitable: When the researchers began the development process, they identified measures that could not be supported with public data. The research team had to compromise by creating measures less detailed than originally conceived, as some data elements were not available in a public statewide dataset. Flexibility in developing measures is also important in order to reflect critical performance issues, and these issues will change over time (Pink, et al).
2. Data quality is a concern: The use of performance measures is dependent upon valid, reliable, and comparable data. In accessing public data, there were some limitations due to missing data or problems with data quality. However, the most frequent feedback the researchers received was poor data quality, not calculation errors. For example, in reviewing data elements retrieved from the California data system, some schools identified inconsistencies with data on teachers. Very proactive schools exercise their right to petition the State to correct the data in the California system. As schools become more reliant on the data and use increases, data quality will likewise increase based on past research.
3. Comparisons are valuable: Comparisons can help to determine if a school is out of line with other schools. For example, once schools compare their performance with other schools, they get a new perspective on their performance and meaningful measures. If they are performing better than other schools, they will want to widely share this information. If they are performing worse than other schools, they may want to do a more rigorous self-evaluation of their performance to determine areas for improvement or they may want to contact other schools doing better to research promising practices.
4. Importing existing measures is valuable: This evaluation system includes three existing measures developed by others. They measure academic performance (API), annual academic progress (AYP) and teacher qualification (TQI) and have been broadly accepted or approved for statewide use. Rather than developing new measures, the API and AYP were accepted as part of the system. These measures have been politically vetted and are widely accepted. The other indices are original and provide additional school performance measures not currently in existence.
5. Involving stakeholders in the process improves the product: Representatives from the education community, including charter school operators, education service providers, and state and district administrators, helped the research team identify data quality problems and methodological issues that could not have been identified by the researchers in isolation. They also shared multiple stakeholder perspectives about the utility of the measures and areas for future exploration. The input of stakeholders in the development of multiple measures facilitated four lessons:
  - a. Their involvement provided broad support for the evaluation system;
  - b. Their involvement ensured useful suggestions and sharing of experiences/expertise that improved the development of the performance measures;

- c. Their engagement made them more aware of the strengths and limitations of the existing accountability system and options available; and,
  - d. Their involvement reinforced the need for a Web-based performance evaluation system for schools.
6. Information is political: While multiple measures can be considered progress toward new levels of accountability, one must have caution. Given that performance information on schools has been used in political venues, if measures are not interpreted correctly or are used inappropriately, a well-intentioned evaluation system could actually hinder improvement. In addition, the availability of comparative information generates a set of new and often unforeseen uses of the data that have political consequences not always anticipated.

## Conclusion

This study is important to educational practitioners for a number of reasons. First, this development was initiated with a group of charter schools because they are public schools granted flexibility from educational regulations in exchange for increased accountability. Leaders of these schools are acutely aware of the need to create multiple measures of performance in a system that can help them track their performance in multiple areas, identify performance areas for improvement, inform their data-driven decision making, “tell their story” to external stakeholders, and complete their charter renewal petitions with relevant data. This research significantly assists charter schools in these educational efforts. Secondly, although the performance evaluation system development efforts focused on charter schools in a single state, the system is nonetheless generalizable to all public schools – charter and non-charter public schools and in other contexts— because public data required of all public schools are used.

The underlying argument of this article rests on the idea that multiple measures of both educational inputs and outcomes are needed to address public accountability for quality schools as well as the school’s internal accountability to improve performance. As the flow of information improves, so too will the quality of educational services. Schools will have data to track their performance and to improve the quality and sustainability of their programs. Further, as the availability of data becomes more widespread, stakeholders will become better educated and engaged in the educational process, an outcome of great educational significance. It is hoped that researchers, practitioners, and policymakers who are engaged in performance measurement and accountability endeavors will draw from the development process that was applied here.

By accomplishing the two objectives – 1) integrating data used primarily for compliance purposes into practical measures for improving school quality and sustainability, and 2) broadening measures of school accountability from a single performance measure to multiple measures– the performance evaluation system is better able to stimulate school improvement efforts and to provide comparative data to facilitate information sharing among peer schools. With more measures to assess performance, greater levels of accountability and school performance can also be achieved.

## References

- Bensimon, E. M. (2004). The diversity scorecard: A learning approach to institutional change. Change. [www.eric.ed.gov](http://www.eric.ed.gov).
- Bensimon, E.M. (2005). Closing the achievement gap in higher education: An organizational learning perspective. *New Directions for Higher Education*, 131, 99-111.
- Brown, R.S. (1999). Creating school accountability reports. *The School Administrator*. November, 12-17.
- DeBray, E. H., McDermott, K. A., & Wohlstetter, P. (2005). Introduction to the Special Issue on Federalism Reconsidered: The Case of the No Child Left Behind Act. *Peabody Journal of Education*, 80(2), 1-18.
- Kaplan, R. S. & Norton, D. P. (1992). The balanced scorecard: Measures that drive performance. *Harvard Business Review*, 70, 64-72.
- Kaplan, R. S. & Norton, D. P. (1993). Putting the balanced scorecard to work. *Harvard Business Review*, 71, 74-87.
- Meyer, C. (1994). How the right measures help teams excel. *Harvard Business Review*, 72, 45-54.
- Pink, G.H., McKillop, I., Schraa, E.G., Preyra, C., Montgomery, C., Baker, G.R. (2001) "Creating a balanced scorecard for a hospital system." *Journal of Health Care Finance*, Spring 2001.
- Quinlivan, D. (2000). Rescaling the balanced scorecard for local government. *Australian Journal of Public Administration*, 59(4), 36-41.
- Stewart, A. C., & Carpenter-Hubin, J. (2001). The balanced scorecard: Beyond reports and rankings. *Planning for Higher Education*. Winter, 27-42.
- Storey, A. (2002). Performance management in schools: Could the Balanced Scorecard help? *School Leadership & Management*, 22(3), 321-338.
- Urrutia, I. & Eriksen, S.D. (2005). Application of the balanced scorecard in Spanish private health-care management. *Measuring Business Excellence*, 9(4), 16-26.